Description

A Method and Apparatus for Rate Distortion Optimization Based Rate Control

BACKGROUND OF THE INVENTION:

Advanced video coding techniques are important for multimedia storage and transmission. For this reason, many video coding standards have been standardized. H.264 is the latest video coding standard. H.264/AVC standard jointly developed by ISO and ITU-T—Joint Video Team (JVT), also known as MPEG-4 Part 10 and H.264 in the H.26x serial standards, has substantially outperformed the previous video coding standards by utilizing a variety of temporal and spatial predictions. Rate control is an important technique although it does not belong to the normative part in video coding standards. However, without rate control any video coding scheme would be practically useless in many applications because the client buffer may often under-flow and over-flow when a channel used to deliver the compressed stream is of constant bandwidth. Therefore, every video coding standard has its own rate control technique, for example, TM5 for MPEG-2 and TMN8 for H.263.

RDO is one of important video coding techniques. It is used to select optimal motion vectors an optimal coding mode for every macroblock. Yet the RDO used in H.264 test model makes it difficult to adopt the existing rate control techniques. Because rate control usually requires a pre-determined set of motion vectors and coding modes to select the quantization parameter, whereas RDO requires a pre-determined quantization parameter to select motion vectors and coding modes. On the other hand, as the complexity ratio between coded frame, the bit allocation model and adaptive quantization scheme should also be improved. The invention is a method and apparatus for rate distortion optimization based rate control. The invention can be used for video streaming, transmission, and storage coding.

SUMMARY OF THE INVENTION:

The invention is to provide a method and apparatus of rate control for a video encoder, in which rate distortion optimization technique is used to improve coding efficiency.

As shown in Figure 2, a rate distortion optimization based rate control implementation includes following modules: JVT processing module, rate distortion optimization based macroblock mode selection module, virtual buffer, and global complexity estimation module.

JVT processing module receives the input frame data, and it is connected with RDO mode selection module, virtual buffer module and global complexity estimation module;

RDO mode selection module is connected with virtual buffer and global complexity estimation module. It receives the input signal from JVT processing module, and processes it based on the virtual buffer module and global complexity module status. In the last, the output signal is sent back to JVT processing module, JVT module will output the final coded macroblock.

Before coding a GOP, does bit allocation for the pictures in the GOP with the average picture size; The average picture size is calculated as:

$R/F = R \div F$, here, R is the target bit rate. F is the picture rate. R/F is the average picture size.

The bit allocation adjustment in the coded GOP is shown as follows:

$$T_b = \max\left\{\frac{R}{N_b + \dfrac{N_p K_b Xp}{K_p X_b}}, \frac{bit\_rate}{8 \times picture\_rate}\right\}$$

$$T_p = \max\left\{\frac{R}{N_p + \dfrac{N_b K_p X_b}{K_b X_p}}, \frac{bit\_rate}{8 \times picture\_rate}\right\}$$

$$T_i = \max\left\{\frac{R}{1 + \dfrac{N_p X_p}{K_p X_i} + \dfrac{N_b X_b}{K_b X_i}}, \frac{bit\_rate}{8 \times picture\_rate}\right\}$$

here, T_I, T_p and T_b is the bits allocated to the I, P or B frame respectively. N_i、 N_p and N_b is the remained none coded I, P or B frames in the GOP respectively. X_i、X_p and X_b is the global complexity estimation for the I, P or B frame respectively and is defined as the multiplier between coded bits and average quantization parameter for the frame.

bit_rate is the target bit rate. *picture_rate* is the frame rate.

K_p and K_b are constants. K_p, K_b means the complexity ration between P, B frame and I frame respectively.

R is the remained bits for the GOP, and after coding a picture is updated as follows:

$R = R\text{-}S_{i,p,b}$

$S_{i,p,b}$ is the coded bits for the current frame.

Before coding a GOP, the remaining bits for the current GOP is initialized as follows:

$R = G + R_{prev}$

$G = bit\_rate \times N \div picture\_rate$

here, R is the remained bits for the current GOP.

N is the number of frames in current GOP.

G is the number of bits for a GOP.

R_{prev} is the remained bits for the previous GOP. For the first GOP, R_{prev}=0.

X_i, X_p and X_b are initialized as:

X_i=a $\times bit\_rate$

X_p=b $\times bit\_rate$

X_b=c×*bit_rate*

here a, b and c are constants.

bit_rate is the target bitrate.

Does the mode selection while using the quantization parameter of previous macroblock as a prediction value for the current macroblock. The mode minimizes the following expression is selected as the initial coding mode for the current macroblock:

$$D(s,c,MODE \mid QP) + \lambda_{MODE} R(s,c,MODE \mid QP)$$

here, s is the luma value of the original macroblock. c is the luma value of the reconstructed macroblock. λ_{MODE} is the lagrangian constant.

For I/P frame, $\lambda_{MODE} = 0.85 \times 2^{Q_{m-1}/3}$;

For B frame, $\lambda_{MODE} = 4 \times 0.85 \times 2^{Q_{m-1}/3}$.

D(s,c,MODE|QP) is used to evaluate the distortion of the current macroblock after it is coded.

R(s,c,MODE|QP) is the bits used to code the macroblock with mode *MODE*.

QP is the quantization parameter for current macroblock.

for motion estimation in P or B frame, the motion vectors minimizes following expression are selected as the motion vectors for the current macroblock:

$$J(\mathrm{m}, \lambda_{MOTION}) = SA(T)D(s, c(\mathrm{m})) + \lambda_{MOTION} R(\mathrm{m} - \mathrm{p})$$

here, *D(s,c(m))* is used to evaluate the distortion from motion compensation.

SA（T）D is sum of the absolute difference (or after Hadmard transform) for the macroblock.

R(m-p) is the bits used to code the motion vector.

s is the luma value of the current macroblock in the original frame.

c is the luma value in reference picture.

m is the motion vector.

p is the predicted motion vector.

λ_{MOTION} is the lagrangian constant and $\lambda_{MOTION} = \sqrt{\lambda_{MODE}}$.

λ_{MODE} is the lagrangian constant.

After the first rate distortion mode selection, the output of RDO mode selection module is

sent to JVT processing module. A new quantization parameter will be calculated by the JVT processing module. The quantization parameter is adjusted according to macroblock activity.

After first rate distortion mode selection, the sum of the absolute difference is used as the macroblock activity estimation. The macroblock activity is calculated as:

$$act_m = \sum_{i,j} |s(i,j) - c(i,j)| \quad N\_act_m = \frac{(2 \times act_j) + avg\_act}{act_j + (2 \times avg\_act)}$$

here, i is the horizontal position of the pixel in the current macroblock. j is the vertical position of the pixel in the current macroblock. N_act_m is the activity of the current macroblock. $s(i,j)$ is the luma value of the original pixel(i,j), $c(i,j)$ is the prediction value of pixel(i,j). avg_act is the average act_m in the previous coded picture which is coded with the same type as current picture. act_m is the sum of the absolute difference after motion compensation or intra prediction.

When coding the first frame, the virtual buffer occupancy is initialized with:

$$d_0^i = 10 \times r / 31$$

$$d_0^p = K_p \times d_0^i$$

$$d_0^b = K_b \times d_0^i$$

here r is the virtual buffer size; d_0^i, d_0^{p}, and d_0^b is the initial virtual buffer occupancy for i, p, or b frame. K_p is the complexity ration between I, P frame; K_b is the complexity ratio between I,B frame.

The RDO based rate control also includes a second RDO mode selection, after quantization parameter decision for the current macroblock. That is to say, the decided quantization parameter for the current macroblock will be used to RDO mode selection again. The mode which minimizes the following expression will be selected as coding mode for the current macroblock:

here, s is the luma value of the original macroblock. c is the luma value of the reconstructed macroblock. λ_{MODE} is the lagrangian constant.

For I/P frame, $\lambda_{MODE} = 0.85 \times 2^{Q_{m-1}/3}$;

For B frame, $\lambda_{MODE} = 4 \times 0.85 \times 2^{Q_{m-1}/3}$.

$D(s,c,MODE|QP)$ is used to evaluate the distortion of the current macroblock after it is

coded.

R(s,c,MODE|QP) is the bits used to code the macroblock with mode *MODE*.

QP is the quantization parameter for current macroblock.

Quantization parameter from JVT processing module is sent back to JVT processing module, the macroblock is coded by JVT processing module and output.

Based on above modules, the drawbacks of traditional rate control schemes are removed. As RDO and rate control are considered together, the RDO based video coding can reach accurate target bitrate control while with good performance.